

L'ANALYSE DE SURVIE

Spécialisé dans le traitement statistique des études cliniques,
le Département Biométrie de FOVEA vous propose
au travers de ce fascicule quelques notions simples
concernant l'analyse de survie.



1

Introduction

Dans le cadre de l'analyse d'un essai clinique, on peut être amené à étudier **des données dites "censurées"**. Ce sont des données concernant des événements qui, lorsqu'ils se produisent, n'apparaissent qu'**une seule fois au cours de l'essai** : décès, guérison, 1^{ère} rechute, 1^{ère} apparition d'un événement indésirable, ...

Relever uniquement la fréquence des événements étudiés peut se révéler insuffisant. Il peut être également important de connaître dans quel délai ces événements sont susceptibles de survenir.

Les méthodes d'analyse de survie permettent d'associer la **fréquence** et le **déla** **survenue de l'événement étudié**. Ces méthodes doivent leur qualificatif au fait qu'elles ont été utilisées d'abord en cancérologie où la survie des patients, c'est-à-dire le temps s'écoulant entre le début du traitement et la survenue du décès, était un des critères d'efficacité du traitement.

l'observation du sujet. L'estimation tient compte des sujets incomplètement suivis.

n Date de point

C'est la date à laquelle on décide d'analyser les résultats.

n Date d'origine

C'est le point de départ à partir duquel le sujet est observé. Ce peut être la date de diagnostic, la date de début du traitement, la date de randomisation, ...

Le point de départ doit être le même pour tous les sujets, mais la date bien sûr varie d'un sujet à l'autre.

n Date des dernières nouvelles

C'est la date à laquelle on a eu pour la dernière fois des nouvelles du sujet :

- ainsi, ce peut être **la date du décès** ou **la date de survenue de l'événement étudié** : guérison, 1^{ère} rechute, 1^{ère} apparition d'un événement indésirable, ...,
- mais ce peut-être aussi **la date de la dernière consultation** si le sujet n'a pas présenté l'événement étudié (décès, guérison, ...) ou si le sujet est perdu de vue.

n Etat aux dernières nouvelles

Aux dernières nouvelles, le sujet est décédé ou vivant, est guéri ou non guéri, a rechuté ou n'a pas rechuté, a présenté un événement indésirable de tel ou tel type ou non, ...

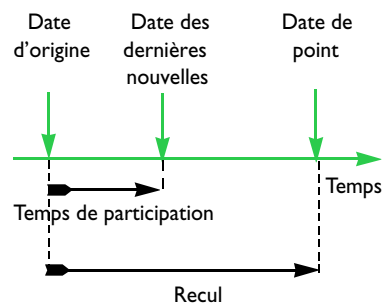
n Recul

C'est le délai écoulé entre la date d'origine et la date de point.

n Temps de participation

C'est le temps écoulé entre la date d'origine et :

- **la date de survenue de l'événement**,
- ou **la date des dernières nouvelles** si le sujet est perdu de vue,
- ou **la date de point** si le sujet est présent pendant toute la durée de l'observation sans que l'événement se soit produit pendant cette période ; dans ce cas, les données sont dites censurées à droite.



n Médiane de survie

C'est le délai à l'issue duquel 50 % des sujets sont encore en vie (ou 50 % des sujets sont guéris, ...).

3

Construction d'une courbe de vie

n Postulat

Si tous les patients d'un essai étaient décédés à la fin de l'essai, il serait aisé de déterminer la proportion de patients encore en vie après tel ou tel délai, calculée à partir d'un point de départ prédéterminé.

On pourrait alors construire un graphique simple joignant

2

Définitions

n Courbe de survie

C'est une courbe qui donne une estimation de la proportion de sujets qui seront encore en vie ou n'auront pas présenté le phénomène étudié, passé un certain délai après le début de



des points dont les coordonnées seraient la proportion de patients vivants (*en ordonnée*) et le délai écoulé depuis le point de départ (*en abscisse*).

Ce cas de figure est irréaliste dans la mesure où, heureusement, tous les patients ne meurent pas au cours de l'essai.

Il n'est pas possible d'estimer de façon simple la proportion de patients encore en vie X jours après la date d'origine :

il faut donc raisonner en termes de probabilités.

Pour calculer un **taux de survie**, il faut appliquer une règle de base simple en apparence mais néanmoins très importante.

On estime que pour survivre une année entière, un patient doit survivre chacun des 365 jours qui composent l'année. C'est-à-dire que la probabilité d'être en vie au 365^{ème} jour est égale à la probabilité d'être en vie au 364^{ème} jour multipliée par la probabilité conditionnelle d'être en vie au 365^{ème} jour, sachant que l'on était vivant le 364^{ème} jour.

Ainsi, la probabilité d'être en vie 1 an après la date d'origine est donnée par la formule :

$$P = P_1 \times P_2 \times P_3 \times \dots \times P_{364} \times P_{365}$$

- P_1 = Probabilité de survivre 1 jour au moins après la date d'origine
- P_2 = Probabilité de survivre le 2^{ème} jour sachant que l'on a survécu le 1^{er} jour
- P_3 = Probabilité de survivre le 3^{ème} jour sachant que l'on a survécu les 2 premiers jours
- ...
- P_{365} = Probabilité de survivre le 365^{ème} jour sachant que

l'on a survécu les 364 jours précédents

Les probabilités conditionnelles correspondent au rapport du nombre de patients vivants à la fin du jour J sur le nombre de patients vivants au début du jour J.

Par exemple, la probabilité P_{365} est égale au rapport du nombre de patients vivants à la fin du 365^{ème} jour sur le nombre de patients vivants au début du 365^{ème} jour (*patients exposés au risque de décéder*). On peut rapprocher ce raisonnement du calcul de la probabilité de tirer avec 2 pièces de monnaie, à la fois face avec la 1^{ère} pièce et face avec la 2^{ème} pièce :

$$P = 1/2 \times 1/2 = 1/4$$

Raisonnement que l'on peut également énoncer ainsi : "Probabilité de tirer face avec la 2^{ème} pièce sachant que l'on a tiré face avec la 1^{ère} pièce".

n Réalisation pratique

Une courbe de survie s'inscrit dans un quadrant déterminé par 2 axes :

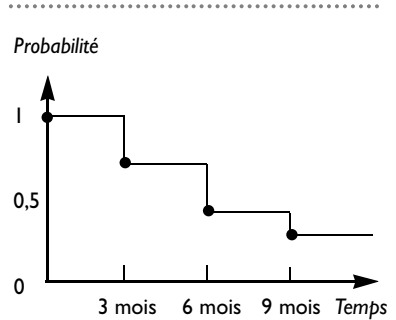
- un axe des ordonnées, étalonné de 0 à 1, représentant les probabilités de survenue de l'événement étudié pour une période déterminée,
- un axe des abscisses étalonné en unités de temps (*jour, mois, ...*).

Chacune des estimations des taux de survie est représentée par un point. Une courbe de survie a l'allure d'un escalier dont la hauteur des marches est déterminée par la différence entre les probabilités de survenue de 2 épisodes successifs de l'événement étu-

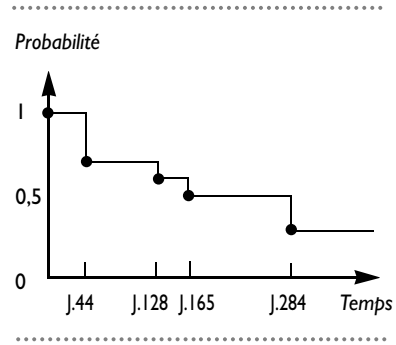
dié (P_A et P_B), et la longueur des marches par le délai écoulé entre la survenue de chacun des 2 épisodes P_A et P_B .

2 méthodes existent pour la construction d'une courbe de survie :

- **La méthode actuarielle** où les périodes de calcul du taux de survie sont prédéterminées et de même durée (*1 mois, 3 mois, 6 mois, ...*). Ainsi, les "marches d'escalier" déterminées par cette méthode sont toutes de longueurs égales. Cette méthode a pour avantage de limiter le nombre de points.



- **La méthode de Kaplan-Meier**, la plus utilisée actuellement, où l'on calcule un taux de survie à chaque survenue d'un épisode de l'événement étudié. On obtient des "marches" de longueurs inégales.



4

Comparaison de plusieurs courbes de survie

On peut comparer 2 taux de survie (*taux de survie de 2 traitements A et B par exemple*) à un moment donné, 6 mois par exemple, en appliquant la formule :

$$z = \frac{S_A - S_B}{\sqrt{\text{var } S_A + \text{var } S_B}}$$

S_A et S_B étant les taux de survie calculés à 6 mois.

Il semble cependant plus pertinent de comparer les courbes de survie dans leur ensemble.

2 méthodes sont principalement employées :

- le test du Logrank,
- le modèle de Cox.

Le test du Logrank ne permet de prendre en compte qu'**un seul facteur**, le modèle de Cox permet d'étudier simultanément **plusieurs facteurs**.

Le test du Logrank a été décrit par Mantel en 1966. Le principe en est le suivant : si pour un jour donné, la moitié des patients observés sont dans le groupe A et la moitié dans le groupe B, alors en moyenne la moitié des décès devrait survenir dans le groupe A et la moitié dans le groupe B, à moins que le traitement A soit beaucoup moins efficace que le traitement B, ou l'inverse. Ainsi, on recense dans chacun des 2 groupes, le nombre observé de décès (O_A et O_B) pendant une période considérée. On détermine ensuite le nombre

estimé de décès que l'on devrait observer si la mortalité était la même dans les 2 groupes (E_A et E_B).

On compare O_A à E_A , O_B à E_B au moyen d'un test du χ^2 dont l'interprétation se lit dans une table pour un degré de liberté (1 ddl). Si l'on compare k courbes de survie, on utilise alors un test du χ^2 avec $(k-1) \text{ ddl}$.

Le test du Logrank permet également de calculer, dans chaque groupe, le **taux relatif de décès** (O/E). Le rapport des taux relatifs de décès de chacun des 2 groupes

$$(X = \frac{O_A/E_A}{O_B/E_B})$$

s'appelle le **risque relatif** (RR). Cela permet de conclure que le risque de décès dans le groupe A est X fois supérieur à celui dans le groupe B ($X = RR$).

Le modèle de Cox (qui sera décrit plus amplement dans un fascicule ultérieur), proche du modèle de régression logistique, est un modèle multivarié qui permet d'exprimer la relation entre le risque instantané (λ_i) de survenue de l'événement étudié et des facteurs de risque exprimés sous la forme de variables X_i qualitatives ou quantitatives.

5

Conclusion

Les méthodes utilisées pour étudier la survie ou tout autre événement survenant de façon "unique" au cours d'un essai clinique permettent :

- d'obtenir une courbe de survie : description graphique des taux de survenue de l'événement étudié dans un ou plusieurs groupes de traitement ;
- de déterminer la probabilité de survenue de l'événement étudié après un certain délai ;
- de comparer la probabilité de survenue de l'événement étudié entre différents groupes de traitement ;
- de mesurer l'influence d'une variable explicative (âge, sexe, sévérité de la maladie, ...) sur la probabilité de survenue de l'événement étudié (*test du Logrank ajusté*) ;
- de stratifier a posteriori, sur la variable explicative, les groupes à comparer et de calculer un nouveau de degré de signification.

Enfin rappelons, que bien que le terme de survie soit le terme consacré pour des raisons historiques, les méthodes d'analyse de survie ne s'appliquent pas seulement à l'étude des décès mais peuvent s'appliquer également à l'étude de tout événement "unique" susceptible d'apparaître au cours d'un essai : 1^{ère} apparition d'un événement indésirable, 1^{er} épisode de rechute, 1^{ère} normalisation d'un critère, ...

La méthode la plus employée, à l'heure actuelle, afin de déterminer une courbe de survie est la méthode Kaplan-Meier.